AB-11 - Paper

Automated Competency Assessment: Potentials and Pitfalls

Ernest M. Paskey
U.S. Office of Personnel Management

Introduction

The rapidly changing world of computer technology has increased the possibilities for new methods of assessment. Voice recognition systems, natural language processing, computer simulations, and the Internet all present opportunities and challenges in assessment. Voice recognition systems offer a new approach in gathering data, natural language processing provides a method for analyzing textual and numeric input, computer simulations can increase the domain of competencies being measured, and the Internet reaches a global audience. This paper describes these emerging technologies, provides examples of how each technology may be used in assessment, and delineates the benefits and limitations of each technology.

Emerging Technologies

Voice Recognition Systems

19990422 020

Voice Recognition Systems (VRS) allow for an efficient and mobile approach to the input of data into computer systems. Instead of using a keyboard and mouse for data entry, users will simple "speak" to the computer. Voice commands, data entry, and voice print security systems are all examples of what can be done with a VRS. Voice commands and data entry are here today. Many of us have "talked" to a computer on the phone when calling directory assistance, or a company's customer service department. VRS's are currently being bundled with word processing software. Users can open, close, and switch between applications using voice commands. Voiceprint, along with fingerprint and retina scanning, is an emerging technology that will provide new security options.

Continuous speech can be dictated by to a computer, but even the best VRS's have a 95-96% success rate. That still leaves a lot of gaps in a 3,000-word document. The immediate hurdle is to develop Continuous Speech Recognition (CSR). CSR involves taking 'natural' speech, i.e., the way people actually talk, not "The...dog...is...brown," and breaking it down into words. This is difficult for a computer, because pauses in speech are not between words, but within them. Advanced CSR algorithms utilize spectrographic analysis, but even that is not the entire solution because the same person never quite pronounces something the same way twice.

Natural Language Processing

Beyond voice recognition is the capability of a computer to understand what the user is saying. How many times have you wanted to tell your computer where to go...and have it understand what you said? The development of Natural Language Processing (NLP)

1

AQI99-07-1409

provides computer systems with the ability to analyze and process textual and numeric data. NLP will be the next big breakthrough in computing. NLP is probably about a ten years away from full development. Two federal agencies applying NLP are the Immigration and Naturalization Service and the Department of Defense (DoD). Both users rely on VRS and NLP to translate from a foreign language to English. DoD, for example, uses a system that allows bomb patrols to communicate with local residents about potential bomb locations.

Typically associated with these applications is a high error rate. The challenge is to have computers not only recognize speech, but understand what is meant in limited contexts. Currently, contextual understanding is limited to within specialized applications. For example, in an accounting application, the computer will recognize "interest" in the context of "money" as opposed to "hobbies" or "tropical vacations." Another example would be telling the computer to search the applicant database for the person with skills A, B, and C who is willing to relocate. The computer responds with a listing of applicants meeting the criteria.

Text Analysis Systems

Many resume-scanning systems have evolved from using simple word counts to the utilization of artificial intelligence and neural network analysis algorithms. These systems attempt to analyze the context in which key words are used. Another use for text analysis systems is in the analysis of writing samples. Done primarily in academic settings, text analysis systems rate writing samples efficiently while reducing the need for human raters, which can be costly.

Text analysis systems are also being utilized in the analysis of open survey responses. There are systems currently available that have the capability to quantify interviewees' verbatim responses to open-ended survey questions. These systems enable researchers to quantify and analyze customer's responses to questions such as: "Why do you shop here?" and "What can we do to improve our service?" In the past, analysis of open-ended survey responses had been limited. A coding staff would have to spend days compiling and coding responses with varying degrees of success. Because of the time and money needed for processing and analyzing open-ended survey responses, many times the responses are simply ignored. Automated text analysis provides a solution to the analysis of open-ended responses that is less resource intensive.

While the accuracy of voice recognition, natural language processing, and text analysis systems is continually improving, there may be a constraint on the level of accuracy obtainable by these systems. The limitation is that it is intrinsically impossible for computers to ever thoroughly understand human language. Even if text input is typed into the computer, and the computer has no problem identifying the words, morphology (word parts), syntax (word order and relation), and semantics (sentence meaning) will be difficult for a computer to determine. Even basic word processing is difficult, e.g., a computer would have trouble distinguishing whether "right" meant "turn right", or "write" as in "I like to write", or as in "you are right." Efforts in language processing have been primarily mathematically based, but language is founded more in liberal arts. The highly developed skill of judgement may never be obtainable by a computer. There

is no such thing as an absolute meaning of any word. Context of speech is everything. A spell checker, for example, can be more accurately described as a spell "guider." One must still be a good speller to find any value in a spell checker. While it is possible to program the computer to figure out the probabilities regarding a particular word, only the human mind can handle the many nuances, ironies, and ambiguities of everyday colloquial language.

Computer Simulations

Realistic job previews, situational judgement exercises, and assessments centers all attempt to replicate, to a certain degree, the actual job. The better we can observe a person performing in the job situation, the better we can predict long term job performance. Computer systems allow for job replication by utilizing multi-media scenario simulation.

Computer simulations have been used for human resource purposes extensively by European countries for the past decade (Bakken, Gould, & Kim, 1992; Funke, 1995; Geilhardt & Mühlbrat, 1995). These simulations have been used to measure cognitive and non-cognitive competencies for selection, promotion, and training on both managerial and non-managerial occupations. Simulation testing in the United States has tended to be limited to performance-based assessment, such as pilot and air traffic controller simulations. Part of the difference in approaches in due to theoretical perspective and legislative climate. Simulations utilized in the European community have rarely been subjected to validity studies (Kleinmann & Straub, 1998). In addition, explorations into what is actually being assessed have had limited results (Funke, 1998). Similar to the work being done in the field of neural network analysis, often the relationship between predictors and job performance is recognized, but identifying specific elements of prediction or explanation of the relationship has been problematic. This vagueness of understanding has limited efforts in the United States in implementation of broad-based computer simulations.

Funke (1998) summarized several advantages and disadvantages of computer-based simulations used for selection purposes:

Advantages

(1) Capability to construct highly complex scenarios that behave dynamically over time.

(2) Capability to economically present complex scenarios.

(3) Capability to quickly compute results.

(4) Capability to present complex scenarios in a standardized manner.

(5) High acceptance from the test taker's point of view.

Disadvantages

1. Simulations can become so complex that even the developer of the system does not know what the correct or best solution is for a given problem.
2. Simulations produce a lot of behavioral data, but the psychological interpretation may be unclear for much of the data.
3. Simulations cannot easily be evaluated with respect to their simulated domain validity, i.e., how representative is the simulation to the actually work situation
4. Results from complex simulations cannot be easily compared from one subject to another because the dynamic situations differ between subjects due to their different interventions.
5. Simulations are low on the social dimension, i.e., the individual works through the assessment without interaction with other people.

Computer simulations can provide for accurate job replication. Providing an environment similar to the actual job can enhance the prediction of an applicant's job performance by observing the applicant's performance in a simulation. For example, automated in-basket exercises are currently being used for predicting managerial performance. Virtual office environments have been created to use in the selection of clerical employees. Video-based situational judgement exercises have been used for years as a method for assessment.

Computer technologies, such as speech recognition, natural language processing, video and audio playback, and artificial intelligence will contribute to the development of virtual reality simulations. Simulations that measure cognitive competencies (e.g., reasoning, reading, and mathematical ability), and non-cognitive competencies (e.g., interpersonal skills, conscientiousness, and leadership), will enhance the ability to assess the "whole person" and have the potential to account for more of the variance in predicting job performance.

Any initiative to utilize simulation technologies in assessment should be evaluated against criteria such as expected increase in validity or utility. The real gain in prediction using simulations is to be made in assessing the non-cognitive competencies. It is often claimed that non-cognitive factors are equally as (or sometimes more) important as general mental ability to job success, but our validity coefficients do not support it. This disparity may be due to the limitations of our current ability to measure the non-cognitive factors. It is often the case that simulations correlate highly with mental ability tests and achieve, at best, similar validity coefficients as traditional written cognitive measures. The question must be asked then, what is the gain in utilizing the simulation? Significant increases in utility may be obtainable by measuring factors previously ignored by past methodologies. This domain is where simulations could flourish; future research in using computer simulations to assess non-cognitive factors will be critical in striving for whole person measurement.

Internet-based Assessment

The vast majority of assessments, whether traditional or leading edge, are deliverable through a global network of computers, the Internet. With the emergence of the Internet as a tool of commerce and communication comes the opportunity to reach thousands of people. An example of an Internet test-delivery model is shown below:

Items, tests, statistics, applicant data, and register information are stored in a central location or server. Test developers, administrators, and hiring officers have varying levels of access to the server. Job announcements, registration information, and applicant data collection originate from the server and are distributed to terminals via the World Wide Web. Tests are delivered from the central server to test locations, which may be secure or open. A secure environment consists of a test site where test takers are monitored, such as a Sylvan Learning Center. An open environment consists of any location where a computer has access to the Internet, such as a university career center, library, or test taker's home. Efficiency in test development and maintenance is enhanced through the use of automated item banks.

Job opportunity listing, resume warehousing, and online applications are all being used today. Selection testing, however, has been limited due primarily to security concerns. Tests administered via the Internet are either non-secure, i.e., rating schedules accessible through any connection to the Internet, or delivered to a secure test site, i.e., Graduate Record Examination (GRE) delivered to a proctored test site. Internet-based testing through a secure test network can be expensive, typically ranging from $55 to $85 per hour. Most secure Internet testing being done is in the area of licensing and certification, where the applicant incurs the cost of testing. Costs associated with open or non-secure Internet testing typically range from $15 to $35 per applicant. Legally, government agencies must incur the cost of civil service examining; therefore, costs can be a major concern when implementing Internet-based testing.

How can the lower cost of open Internet testing be combined with the security of a proctored test site? Three major factors determining the security of the Internet are: 1) software,

2) infrastructure, and 3) user responsibility and acceptance. Software engineers and programmers must continually enhance the security capabilities of the software used. Recent security flaws that have been discovered in web browser and email software are an indication of the vulnerability of current software systems. Hardware architects and government policy must create a better security design for the Internet infrastructure. Although the Internet originated from work done in the Department of Defense, early design decisions were made to make it a non-secure network. The inherent hardware

design of the Internet limits the degree of security obtainable unless large-scale modifications are made to the system and *every* computer connected to the Internet uses the same security precautions. The third factor, user responsibility and acceptance, will have the largest impact. While emerging technologies such as voice printing, fingerprint scanning, and video camera monitoring can increase the security at the point of test delivery, the biggest factor is user acceptance. Although Internet security is vulnerable, an ever-increasing number of people are relying on it for financial and commercial transactions. Comparable to credit card fraud, Internet security breaches are quickly becoming considered a factor in the cost of doing business. The annual death toll on U.S. highways is 42,000, but no one suggests abandoning the road system, yet many critics oppose the use of the Internet for examining. In the future, Internet security compromises will be become an accepted "necessary evil" when utilizing the network. This is not to say one should ignore the compromises associated with Internet-based test delivery, rather the testing needs and climate should be evaluated against the risks.

Summary

This paper has discussed a few of the emerging technologies that have potential for implementation in the area of assessment. Other technologies such as artificial intelligence, computer adaptive testing, and neural network analysis also have potential for contributing to better practices in human resource management. While many of the technologies are still far from full development, and many limitations exist, organizations can recognize great gains utilizing these tools as they stand today. Voice Recognition Systems and Natural Language Processing provide new approaches to data collection. Implementation of text analysis tools can decrease the number of human raters needed and increase the efficiency of the rating process. Development of computer simulations can expand the number of predictive factors measured in an assessment. Delivering assessments via the Internet can increase the applicant pool and efficiency of the process. Incorporating new technology provides for great opportunity in both research and application of novel, efficient, and effective approaches to assessment.

References

Bakken, B., Gould, J. and Kim, D. (1992) Experimentation in learning organizations: A management flight simulator approach. *European Journal of Operational Research*, **59**, 167-182.

Funke, J. (1998) Computer-based testing and training with scenarios from complex problem-solving research: advantages and disadvantages. *International Journal of Selection and Assessment*, **6**, 90-96.

Funke, U. (1995) Using complex problem solving tasks in personnel selection and training. In Frensch, P.A. and Funke, J. (eds.), Complex problem solving: the European Perspective, 219-240. Hillsdale, NJ, Lawrence Erlbaum Associates.

Geilhardt, T. and Mühlbradt, T. (eds.) (1995) *Scenarios for personnel and organization management*. Göttingen, Verlag für Angewandte Psychologie.

Kleinmann, M. and Straub, B. (1998) Validity and application of computer-simulated scenarios in personnel assessment. *International Journal of Selection and Assessment*, **6**, 97-106.

INTERNET DOCUMENT INFORMATION FORM

A . Report Title: Automated Competency Assessment: Potentials and Pitfalls

B. DATE Report Downloaded From the Internet   4/21/99

C. Report's Point of Contact: (Name, Organization, Address, Office Symbol, & Ph #):   Navy Advancement Center
Attn: Dr. Grover Diel (850) 452-1815
Pensacola, FL

D. Currently Applicable Classification Level: Unclassified

E. Distribution Statement A: Approved for Public Release

F. The foregoing information was compiled and provided by:
DTIC-OCA, Initials: VM_ **Preparation Date**: 4/21/99__,

The foregoing information should exactly correspond to the Title, Report Number, and the Date on the accompanying report document. If there are mismatches, or other questions, contact the above OCA Representative for resolution.